AFRL-AFOSR-JP-TR-2016-0028

Data Mining Research for Information Security

Kevin Barton
Texas A&M University-San Antonio

01/29/2016
Final Report

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/ IOA
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
01-02-2016	Final	20-05-2014 to 19-05-2015

4. TITLE AND SUBTITLE
Data Mining Research for Information Security

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA2386-14-1-4052

5c. PROGRAM ELEMENT NUMBER
61102F

6. AUTHOR(S)
Kevin Barton

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Texas A&M University-San Antonio
1 UNIVERSITY WAY
SAN ANTONIO, TX 78224-3134 US

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AOARD
UNIT 45002
APO AP 96338-5002

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/AFOSR IOA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
A DISTRIBUTION UNLIMITED: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Machine-learning and Ontology Assisted Assessment of Research Trends (MOAART) advances machine learning by developing and testing an ontology-based inferencing engine to filter, sort and rank abstracts in specific research areas. The MOAART reports emerging global research trends in a given research area and defines a relevant ontology to current concepts of interest to the Asian Office of Aerospace Research and Development (AOARD) management team. The MOAART also extends visualization tools to support an ontology framework as well as organization of final results. The research employed supervised ensemble machine learning methods to develop an ontology-based inferencing engine capable of filtering, sorting, and ranking research of most interest to the AOARD team.

15. SUBJECT TERMS
Cyber, AOARD

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Kevin Barton
Unclassified	Unclassified	Unclassified	SAR		19b. TELEPHONE NUMBER *(Include area code)* 210-434-6711

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Final Report

"Development of Automated Malware Analysis Framework"

2015/9/13

Keiji Takeda
Keio University
keiji@sfc.keio.ac.jp

1 Overview of the Result

The research project was conducted from June 2014 to June 2015 by the malicious software (malware) research team in Keio University. The out come of the research includes development of a new method for identification of malware, a new method to monitor behavior of malware binary program and platform to analyze malware using both static analysis approach and dynamic analysis approach.

The goal of the project was to develop automated system to analyze malware with minimum human interaction. The developed technologies through this research project are applied to the platform developed and provided semi-automated functionality. Proposed methods are verified their performance against actual malware on the developed platform. Two research papers were published in academic conferences.

2 Achievements

Through this research activity following outcomes were achieved.

- Design and specification for automated malware analysis system were determined.
- Two algorithms for malware analysis were proposed.
- Developed technologies were verified with actual malware.
- API/interfaces for malware analysis system were defined.

The system developed is a set of servers including honeypot to collect malware and modules to conduct static analysis of malware and benign binary program on windows platform and management functions of connected hardware platform and virtual machines.

The platform is developed as an integrated system that provides required

function for analyst to collect, store, conduct static analysis of binary program, conduct dynamic analysis on binary program then manage obtained data. The system enhance capability of security analyst by providing semi-automated malware analysis functions.

3 Research Process

This research has been conducted through following steps.

a. Specification and design of the prototype components.
b. Procurements for prototype components.
c. Implementation of prototype components.
d. Preliminary experiments with prototype components.
e. Specification and design of prototype system
f. Procurements and implementation of prototype system
g. Completion of experiments with prototype system

4 Research Result

Findings through the research activities were complied as research papers. Those papers are presented at a research symposium and workshop.

The research paper related to the static analysis part is titled as " Proposal for Techniques to Identify Files to Investigate by Executable File." This research proposed an approach to generate a profile model of executable files to individually evaluate and extract executable files present on a computer, and that might cause malicious behavior. This allows investigators to identify files that can potentially cause malicious behavior for detailed investigation in a relatively short time, even if the investigators have limited technical expertise. In this research, we created a prototype system for the method of classifying executable files on a computer by extracting their characteristic attribute information as metadata, and performing a regression analysis of these metadata characteristics against known executable file types. When the created regression model was applied to actual malware, it was able to evaluate the majority of the malware as software with a low level of

trustworthiness. This allows investigators to reduce the number of files that should be subjected to detailed analysis under the circumstances of an actual computer security incident. Profile generation by regression analysis successfully determined 95.6% of the malware in the sample as executable files other than those associated with vendor applications, which are potentially dangerous.

The title of research paper on dynamic analysis is "Stealth malware analysis using taint propagation on virtual machine monitor." In this paper, a method for dynamically interpreting semantics information using taint propagation and automatically analyzing malware that uses code injection or rootkits. Also a prototype system for evaluation is developed and applied to analyze actual malware. Because the proposed method can accurately analyze even sophisticated malware of the type used in targeted attacks, it can contribute to swift comprehension and elimination of malware behavior.

Overhead on some parts of context changes, discontinuity in taint propagation, and insufficient countermeasures for malware that changes its behavior through commands remained as points for future improvement.

However, with regard to the problems that surround semantic gap, the proposed method solved them by monitoring data structures through taint analysis based on pages. Furthermore, the method realized VMI without agent insertion into guest OS, which was a factor in detection by malware.

5 Future Works

This research has presented a concept of automated malware analysis platform so the implementation of the system is still in its early stage. Some more improvements and examination with other approaches are expected as future research.

6 Reference (Publication)

[1] A Proposal for Techniques to Identify Files to Investigate by Executable File Profiling, Keiji Takeda□ and Kohei Tsuyuki□., Computer Security Symposium 2014 22 - 24 October 2014.

[2] Stealth malware analysis by using taint propagation on virtual machine monitor, Yuma Kurogome and Keiji Takeda., Computer Security Symposium 2014 22 - 24 October 2014.